

Increasing authorship identification through emotional analysis

Ricardo Martins, José João Almeida, Pedro Henriques, Paulo Novais

Algoritmi Centre / Department of Informatics

University of Minho, Braga - Portugal

ricardo.martins@algoritmi.uminho.pt, {jj, prh, pjon}@di.uminho.pt

Keywords: Natural Language Processing, Sentiment Analysis, Machine Learning

Abstract: Writing style is considered the manner how an author expresses his thoughts, influenced by language characteristics of an individual, period, school, or nation. Most of the times, this writing style can identify the author. Yet, one of the most famous examples comes from 1914 in Portuguese literature, with Fernando Pessoa and his heteronyms Alberto Caeiro, Álvaro de Campos and Ricardo Reis, who had completely different writing styles and led people to believe that they were different individuals. So, the discussion about authorship identification already exists along a century. Currently, there are several alternatives to identify authors of text, however, these solutions do not consider the emotion contained in the text as source of information in the writing style. This paper is about a process to analyse the emotion contained in social media messages as Facebook^a in order to identify the author's emotional profile and use it to improve the ability to predict the authors of the messages. Using preprocessing techniques, lexicon-based approaches and machine learning, we achieved an authorship identification improvement around 5% in the whole dataset and more than 50% in specific authors, when considering the emotional profile on the writing style.

^a<http://www.facebook.com>

1 Introduction

Since Barack Obama's election, the politicians are using social media to have a direct contact with their voters and increase its credibility with their posts and comments. On the other hand, this direct channel enables a correct perception by the voters about the politics, creating opinions about the subjects they consider important. This phenomenon is increasingly turning politicians into digital influencers. So, the way as they communicate in social media can be considered their "personal signature"; so their worries about the way how they can be interpreted are equally important.

With massive information from social media, the digital influencers and their legion of followers validate, reinforce and amplify news, many times faked. As the main objective of these individuals is be "liked, loved and shared", it is very important to choose correctly the words contained into their texts, in order to maximize the sentiment raised up in the readers.

So, the emotional characteristics contained in the messages make up an "emotional profile" about the author and which, along with the words used in the text, helps to determine the message's author profile while writing.

For example, the following posts are from different authors and deal the same theme - the Paris Climate Agreement - however, the writing styles are different and arouse different emotions. While the first balances positive and negative words in the text, the second mostly uses words with negative emotions:

"Today marks a crucial step forward in the fight against climate change, as the historic Paris Climate Agreement officially enters into force. Let's keep pushing for progress" (Barack Obama);

"I'm optimistic we can stop climate change and help those who are being hurt the most by it—all while meeting the world's energy needs" (Bill Gates).

In this paper, we present an approach using the author emotional profile in order to improve the authorship identification.

Section 2, introduces the concept of emotion and presents some theories for emotion representation and analysis. Section 3 presents some work in this area to detect emotion from social media, while Section 4 describes our proposal explaining the steps used in our analysis, and discusses the results obtained from a set of tests performed, and finally, the paper ends in Section 5 with the conclusion and future work.

2 Emotion theories

Historically, several models have been created in an attempt to systematize the emergence of emotions, their associated behaviours and discuss how emotions are elicited in our cognitive system. The main research theories are: discrete, dimensional and appraisal theories.

Discrete emotional theories propose the existence of basic emotions (*happiness, anger, sadness, surprise, disgust, and fear*, for instance) that are universally displayed and recognized. Discrete models group emotions into categories and assumes that they are independent. In the literature, among the discrete models, a well known model is the so-called basic emotions, proposed by Ekman [3]. This model proposes the existence of six basic emotions: happiness, sadness, fear, anger, surprise and disgust. One of the main advantages of discrete models is that, through psychophysical experiments, the perception of emotions by human beings is discrete.

In opposition to this view, **dimensional theories** try to explain emotions in terms of two or three dimensions. The most frequent dimensional characterization of emotions uses two dimensions: arousal and valence. Valence is related to a positive or negative evaluation and is associated to the feeling state of pleasure (vs displeasure). Arousal reflects the general degree of intensity felt. Low arousal is associated to less energy and high arousal with more energy. However, using this two-dimensional is difficult to differentiate emotions that share the same values of valence and arousal, as *anger* and *fear*. For this reason, it is common to add a third dimension to support this differentiation. According to Leventhal [5] “the third view emphasises the distinct component of emotions, and is often termed the componential view”.

Emotional cognitive psychologists focus their studies mainly on the **appraisal process**. According to Scherer [9], the central idea is that emotions are triggered and differentiated by a subjective analysis of an event, situation or object. This cognitive assessment performed personally is called appraisal. For instance, Paul and John are watching a basketball game where their favourite teams are playing. John’s team wins (event). Paul’s appraisal is that an undesirable event happened: his team lost. He is sad. For John the situation’s appraisal is that the event is desirable and he is happy. So, emotion and reason are not disconnected. In fact, emotions require cognitive processes to generate or retrieve preferences and meanings. Emotions are triggered by personal interpretation of the annoying or cheerful aspects of an event, the appraisal. And is the appraisal, a cognitive process, that triggers the emotions.

Despite of different theories, they have in common the sense of positive and negative emotions. According

to Cambria [2], “polarity detection is a popular Natural Language Processing (NLP) task focusing on the binary classification of snippets of text into either positive or negative”. In other words, polarities are the scores associated to positive, negative and neutral parts of a sentence.

3 Related work

Despite the vast amount of works using sentiment analysis, none of them considers the author’s emotional profile as a component of the writing style for authorship identification. So, each work cited below has inspired partially our work as will be mentioned.

The usage of emotions in social media was inspired by the work of Schwartz et al [10], which predicts the individual well-being, as measured by a life satisfaction scale, through the language people used on social media. This is made using randomly selected posts from Facebook and a lexicon-based approach to identify the text words polarities.

Moreover, Baldoni et al [1] have presented an another interesting work involving lexicons and ontologies to extract emotions including sadness, happiness, surprise, fear and anger, which contributed in the emotional profile creation.

Yet, the framework developed for authorship identifying based in online messages presented by Zheng et al [11] that considers features as syntactic, lexical, structural and content-specific contributed in the use of machine learning techniques in order to predict authorship.

4 Data analysis

In order to predict the authors of a post based on the emotion contained in text, 2100 Facebook posts were collected from 8 different authors of different areas, as presented in Table 1. All data was collected at the same time span, reducing temporal situations interference in the text emotions. In order to compare all information, the posts were manually labelled into 2 categories: politicians and non-politicians.

The task of predict the author of a text is composed of several intermediates steps. First, it was needed some preprocessing tasks in order to reduce data size by removing unnecessary text from the original message.

Preprocessing is a very important step in text mining processes and applications. It is the first step not only for text mining approaches but also in data mining.

Table 1: Posts authors

Author	Area	Category
Barack Obama	Politics	Politician
Bill Gates	Business	Non-Politician
Donald Trump	Business	Non-Politician
Hillary Clinton	Politics	Politician
Jeremy Corbyn	Politics	Politician
Leonardo Di Caprio	Entertainment	Non-Politician
Magic Johnson	Sports	Non-Politician
Theresa May	Politics	Politician

There are several preprocessing techniques useful in order to extract information from text, and their usage is according to the characteristics of the information desired. Despite of some techniques were created in data mining, they are useful in text mining approaches, since the same technique can be used for both information extraction, information retrieval, or combined

The preprocessing, after the tokenization, was divided in 3 parallel jobs, as showed in Fig. 1: Part of Speech Tagging (POS-T), Named Entity Recognition (NER) and Stopwords Removal. This strategy was used because both POS-T and NER need the text in the original format, in order to return the correct data from the analysis.

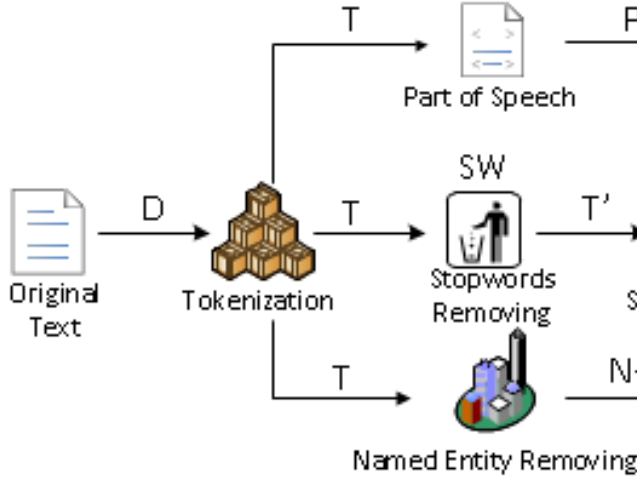


Figure 1: Preprocessing tasks

The POS-T process identifies the text grammatical structure. Concerning text cleaning, only nouns, verbs, adverbs and adjectives were preserved. This is important because only these grammatical categories can bring emotional information. So, in a more formal way, the Tokenization process converts the original text D in a set of tokens $T = \{t_1, t_2, \dots, t_n\}$ where each element contained in T is part of the original document D . Later, the POS-T labels each token with a semantic information. Later, a process collects all nouns,

verbs, adverbs and adjectives in a set P , where $P_T = \{p_{(T,1)}, p_{(T,2)}, \dots, p_{(T,k)}\}$ and $0 \leq k \leq n$ and $P_T \subset T$.

Similarly, NER process identifies names in 3 different categories: “Location”, “Person” and “Organization”. Once identified tokens in one of these categories, they are removed. As a result, a set $N_T = \{n_{(T,1)}, n_{(T,2)}, \dots, n_{(T,j)}\}$ is constructed based on identified word category and where $0 \leq j \leq n$ and $N_T \subset T$. This step is important to be done in parallel with POS because some locations can be confused with nouns (as Long Beach, for instance).

The Stopwords list is a personal predefined set $SW = \{sw_1, sw_2, \dots, sw_y\}$ of words, manually created according to several similar lists available on the internet.

After the 3 preprocessing tasks finish, the result document ST must contain a set of words where $ST = T' \cap P_T \cap N_T$.

Later, in this set ST is applied a stemming process to reduce the words to their word stem in order to consider all inflected words as only one, producing the preprocessed text $PR = \{ST_1, ST_2, \dots, ST_z\}$ ready to be analysed.

For all three tasks - POS-T, NER and Tokenization - the Stanford Core NLP [6] toolkit was used.

An example using a real post from Barack Obama of text preprocessing is presented in Fig. 2.

4.1 Polarity analysis

The first analysis made was aimed at determining the posts polarities. To achieve this objective, after the preprocessing, all sentences contained in PR were compared against EmoLex lexicon [7] in order to identify the positive and negative words contained in the text. This analysis did not take into account the intensity of the polarities neither the emotions.

When comparing the posts’ polarities according to their author’s category (politicians and non-politicians),

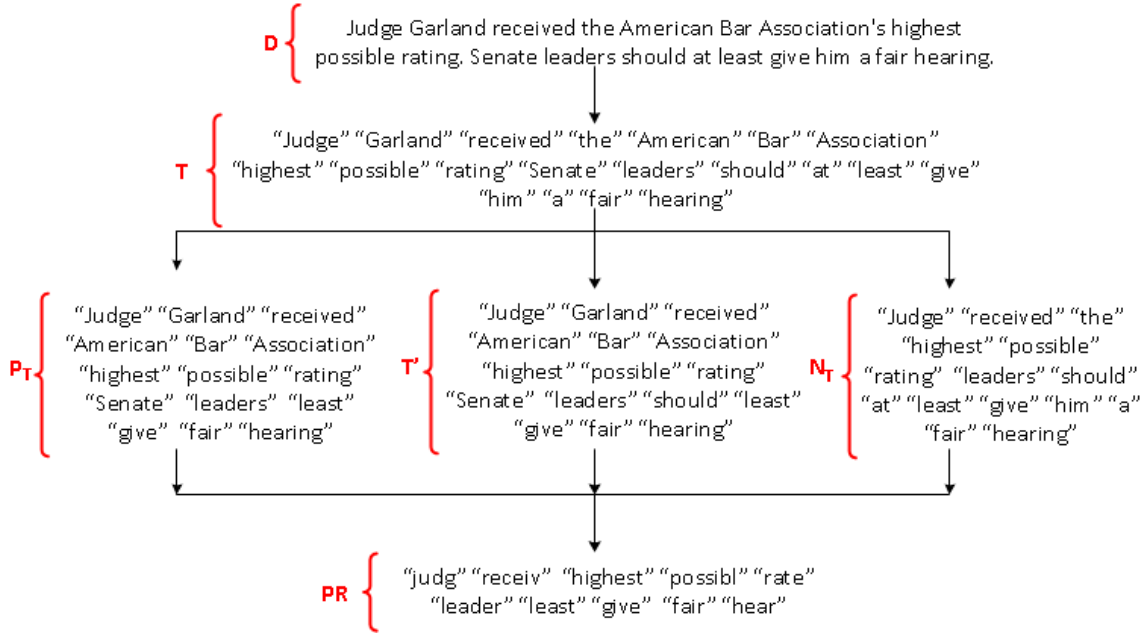


Figure 2: Preprocessing text example

the data did not reveal relevant differences between politician and non-politicians, as showed in the Fig. 3. The same analysis was confirmed using the chi-squared test, where was obtained a value $\chi^2 = 1$, indicating that both polarities data (politicians and non-politicians) are not independent.

However, this interpretation may lead to a wrong understanding about the scenario. When comparing the polarities by author, according to Fig. 4, it is possible to conclude that while politicians tend to have their posts in the same area in a normal distribution, non-politicians tends to be in the extremes, i.e., they are blunter than politicians when expressing through Facebook and indicating that each author has its own "emotional signature" in his posts.

This information is confirmed in Table 2, which presents the positive and negative polarities by author.

Table 2: Polarities by author

Author	Positive	Negative
Barack Obama	0.28	0.13
Bill Gates	0.30	0.11
Donald Trump	0.25	0.16
Hillary Clinton	0.35	0.17
Jeremy Corbyn	0.30	0.13
Leonardo Di Caprio	0.34	0.09
Magic Johnson	0.37	0.06
Theresa May	0.36	0.10

4.2 Lexicon-based emotion analysis

In order to analyse the emotions contained into the text, it was used a lexicon-based approach, which consists in comparing the labelled emotion contained into the EmoLex lexicon with the preprocessed texts described earlier. Using the emotions model proposed by Plutchik [8], where all sentiment is composed of a set of 8 basic emotions (*anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise* and *trust*), all posts where analysed according to this model and a list of emotions in each post was generated, according to Table 3.

Hence, when applying the Person's correlation coefficient (r^2) between polarities and basic emotions, as presented in Table 4, it is possible to point which emotions are related with polarities.

In a scale ranging from -1 to 1, emotions related with a high r^2 value indicates a strong relation with the polarity (as *Anger* and negative polarity), while high negative r^2 values indicates a strong inverse relationship (as *Fear* and positive polarity). In our approach, ambiguous emotions are classified when the standard deviation for r^2 polarity's value is less than 10% range (i.e. 0.2).

In summary, positive and negative emotions are important to describe the author's emotional pattern, while the neutral emotions do not have significant contribution to achieve this objective. the emotions classified in text according to polarities are:

- Positive polarity - Joy;
- Negative polarity - Anger, Fear, Sadness;

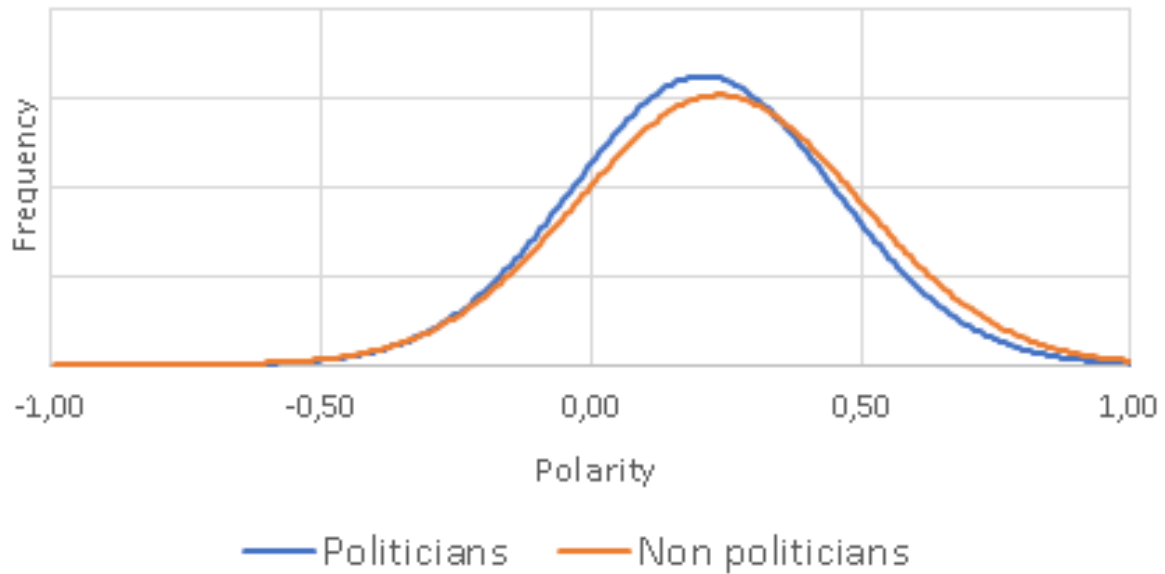


Figure 3: Polarities distribution by category

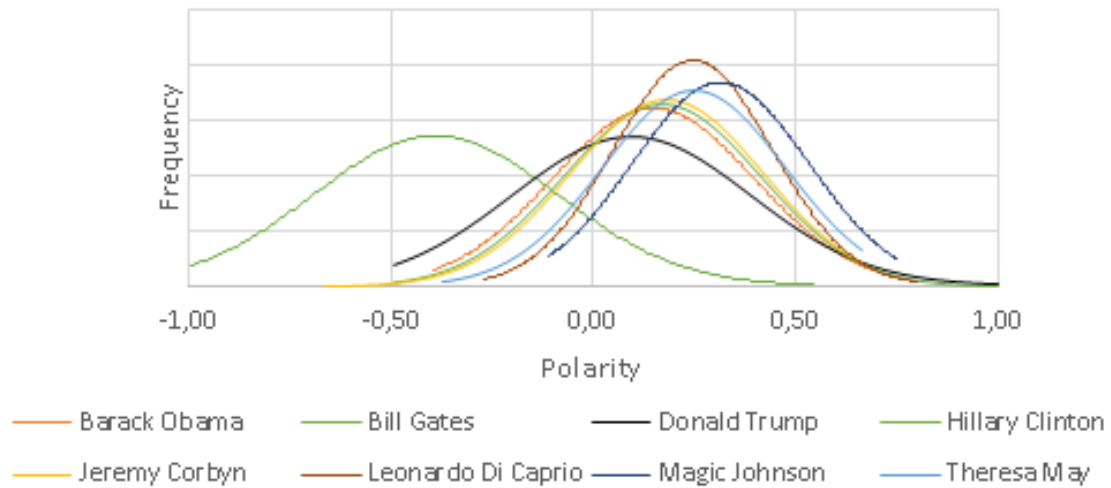


Figure 4: Polarities distribution by author

Table 3: Basic emotions average per author

Author	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
Barack Obama	0.08	0.15	0.03	0.10	0.13	0.05	0.05	0.21
Bill Gates	0.06	0.14	0.04	0.08	0.15	0.06	0.06	0.14
Donald Trump	0.06	0.12	0.02	0.09	0.12	0.10	0.04	0.16
Hillary Clinton	0.14	0.26	0.02	0.07	0.22	0.15	0.12	0.30
Jeremy Corbyn	0.08	0.16	0.03	0.08	0.09	0.08	0.06	0.23
Leonardo Di Caprio	0.04	0.11	0.01	0.07	0.09	0.03	0.03	0.16
Magic Johnson	0.03	0.19	0.03	0.05	0.21	0.04	0.07	0.21
Theresa May	0.06	0.17	0.02	0.06	0.14	0.07	0.07	0.22

- Ambiguous polarity - Anticipation, Disgust, Surprise, Trust.

4.3 Machine learning-based emotion analysis

Once identified the average of each emotion from author, the next analysis was to identify the emotional pat-

Table 4: Correlation between polarities and emotions

Polarity	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
Positive	-0.10	0.49	-0.26	-0.90	0.48	-0.22	0.44	0.40
Negative	0.83	0.27	-0.08	0.60	0.01	0.89	0.34	0.37

tern of the author. To achieve this, it was used an approach based on machine learning (ML) techniques. The first attempt was aimed at identifying the lowest prediction rate author. For this, it was used the same messages in their original state, i.e. with no preprocessing and only the authors identification in a ML approach. Once this information was obtained only by original texts with no preprocessing, this value can be considered the lowest acceptable value, and, in case of decreasing this rate, it may be interpreted as a negative influence of preprocessed texts in the authors prediction. In our initial tests, the best rate was presented by a SVM implementation through Weka [4] and a 10-fold cross validation in the whole dataset, with a correct prediction precision of 82% of when predicting authors.

When the lowest prediction rate was identified, the next step was to classify using the preprocessed information. Using the previous preprocessed texts, polarity values and each basic emotion rate, a new dataset was generated in order to be used in ML process. In our tests, it was used the most relevant algorithms for text classification, as SVM, Naive Bayes, Random Forests, however, using a Naive Bayes Multinomial implementation through Weka and a 10-fold cross validation in the whole dataset, returned a precision of 87.41% of correct predictions when predicting authors. Both results (non-preprocessed and preprocessed) are presented in Table 5.

5 Conclusion

This paper presents a combination of lexicon-based and machine learning approaches to explore the emotions contained in a text through the best practices in sentiment analysis in order to increase the results' accuracy in authorship identification.

Everyone have particular characteristics of expressing themselves, and these personal characteristics can be expressed in their texts.

Once the author's writing style profile is known, by using the emotional information contained into text helps to increase the accuracy on authorship identification. This claiming is based on the successful predictions rate grown from 82% to 87.41% in our tests, besides the values of precision, recall and f-measure which have increased in the majority of the cases, when using

emotional labelled data. This improvement can be interpreted as a very satisfactory result as our proposal.

As future work, it is planned to determine the author's emotional intensity profile, by combining with other text analysis metrics, in order to increase the authorship identification.

Acknowledgements

This work has been supported by COMPETE: POCI-01-0145-FEDER-0070 43 and FCT – Fundação para a Ciência e Tecnologia within the Project Scope UID/CEC/ 00319/2013.

REFERENCES

- [1] Matteo Baldoni, Cristina Baroglio, Viviana Patti, and Paolo Rena. From tags to emotions: Ontology-driven sentiment analysis in the social semantic web. *Intelligenza Artificiale*, 6(1):41–54, 2012.
- [2] Erik Cambria, Amir Hussain, Catherine Havasi, and Chris Eckl. SenticSpace: Visualizing opinions and sentiments in a multi-dimensional vector space. In *KES (4)*, pages 385–393. Springer, 2010.
- [3] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.
- [4] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.
- [5] Howard Leventhal and Klaus Scherer. The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition and emotion*, 1(1):3–28, 1987.
- [6] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60, 2014.
- [7] Saif M. Mohammad and Peter D. Turney. Crowdsourcing a word-emotion association lexicon. 29(3):436–465, 2013.

Table 5: Detailed accuracy results for non-preprocessed and preprocessed texts

Author	Non-preprocessed texts			Preprocessed texts		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Barack Obama	0.933	0.82	0.776	0.907	0.860	0.883
Bill Gates	0.887	0.874	0.836	0.882	0.944	0.912
Donald Trump	0.317	0.465	0.518	0.761	0.556	0.642
Hillary Clinton	0.571	0.686	0.693	0.676	0.762	0.716
Jeremy Corbyn	0.741	0.824	0.814	0.859	0.836	0.847
Leonardo Di Caprio	0.807	0.782	0.734	0.878	0.876	0.877
Magic Johnson	0.867	0.893	0.867	0.945	0.917	0.931
Theresa May	0.495	0.591	0.588	0.649	0.758	0.699

- [8] Robert Plutchik. Emotions: A general psychoevolutionary theory. *Approaches to emotion*, 1984:197–219, 1984.
- [9] Klaus R Scherer, Tim Dalgleish, and Mick Power. Handbook of cognition and emotion. *Handbook of cognition and emotion*, 1999.
- [10] H Andrew Schwartz and Kern Margaret L Sap, Maarten. Predicting individual well-being through the language of social media. In *Biocomputing 2016: Proceedings of the Pacific Symposium*, pages 516–527, 2016.
- [11] Rong Zheng, Jiexun Li, Hsinchun Chen, and Zan Huang. A framework for authorship identification of online messages: Writing-style features and classification techniques. *Journal of the Association for Information Science and Technology*, 57(3):378–393, 2006.